INVESTIGATING GENETIC CONFOUNDING OF THE RELATIONSHIP BETWEEN COLLEGE DEGREE ATTAINMENT AND HEALTH

Adam G. Lilly

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Sociology in the College of Arts and Sciences.

Chapel Hill
2020

Approved by:

Kenneth A. Bollen

Guo

Kathleen Mullan Harris

ABSTRACT

Adam G. Lilly: Investigating Genetic Confounding of the Relationship between College Degree Attainment and Health
(Under the direction of Kenneth A. Bollen and Guang Guo)

Until recently, it was difficult for researchers interested in the relationship between college degree attainment and health to directly test for genetic confounding. This study investigates that question using three separate health outcomes (depression, body mass index (BMI), and self-rated health). To test for genetic confounding, I use a structural equation modeling approach with polygenic scores (PGSs) to compare the effect of a college degree on health across various model specifications. I also conduct analyses using propensity scores to investigate whether PGSs continue to be useful controls over and above a long list of common covariates available in large social science datasets. Results provide evidence for genetic confounding of the relationship between college degree attainment and health when examining BMI and self-rated health, and weaker evidence when examining depression. Propensity scores estimated using widely available covariates seem to account for the entire genetic effect captured by the PGSs.

ACKNOWLEDGEMENTS

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis. I am grateful to the Carolina Population Center for training support (T32 HD091058) and for general support (P2C HD050924).

I would like to thank Elizabeth Lawrence for generously providing the code used to generate the variables used in some of these analyses. I would also like to thank participants at the 2019 Population Association of America Annual Meeting, the 2019 Integrating Genetics and the Social Sciences conference, and in the Thursday Interdisciplinary Workshop at the Carolina Population Center for helpful comments on previous versions of this project. Finally, I would like to thank my advisors, Ken Bollen and Guang Guo, and my other committee member, Kathie Harris, for the all of the advice and feedback they provided during the process of writing this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Add Health	The National Longitudinal Study of Adolescent to Adult Health
BIC	Bayesian information criterion
BMI	Body mass index
CFI	Comparative fit index
GCTA	Genome-wide complex trait analysis
GWAS	Genome-wide association study
IPTW	Inverse probability of treatment weight
LD	Linkage disequilibrium
MLR	Robust maximum likelihood
PGS	Polygenic score
rG	Genetic correlation
RMSEA	Root mean square error of approximation
SEM	Structural equation model
TLI	Tucker-Lewis index
WLSMV	Means and variance adjusted weighted least squares

INTRODUCTION

Social and behavioral scientists have a long history of studying the social determinants of physical and mental health, and have advocated for increased attention to social conditions as fundamental causes of health across all fields of research (Link and Phelan 1995; Phelan, Link, and Tehranifar 2010). One of the most widely studied of these social determinants has been educational attainment (Ross and Wu 1995). While educational attainment is almost always shown to be strongly associated with measures of health, the extent to which this association reflects an underlying causal relationship remains in question (Hermann et al. 2011; Miech and Shanahan 2000; Mirowsky and Ross 2008). One possible omitted variable in studies that consider this relationship is an individual's genetic makeup.

There is some evidence that common genetic variants have effects on both educational attainment and health outcomes (Boardman, Domingue, and Daw 2015). Recent estimates of genetic correlation (rG), which I define below, suggest that some genetic variants positively associated with education are also positively associated with self-rated health, and negatively associated with BMI and depression (Boardman et al. 2015; Harris et al. 2017; Okbay et al. 2016; Wray et al. 2018). These variants could confound the education-health relationship. I use polygenic scores (PGSs) to measure the genetic contributions to education and my health outcomes of interest. PGSs are an index that represents the additive genome-wide influence of individual genetic variants on an outcome.

The purpose of this paper is to assess the importance of genetic confounding for research focused on the education-health relationship by testing whether different PGSs originally devised

either for education or a health outcome have direct effects on both receipt of a college degree and three separate health outcomes. I draw on data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). The effects of the PGSs are first estimated net of adolescent's parental socioeconomic status. Then, I re-estimate the effects of the PGSs after conditioning on a propensity score for college degree attainment that is hypothesized to control for many genetic and non-genetic confounders.

I first briefly review the literature on the relationship between education and health and efforts to better understand the causal association between the two. I also review the concept of genetic correlation (rG) and explain when it could result in genetic confounding. I then describe the models I will use to test whether or not genetic confounding should be a conern. Finally, I estimate multiple specifications of three models examining the relationship between receipt of a college degree and the three health outcomes of BMI, depression, and self-rated health. Results provide evidence for genetic confounding of the relationship between college degree attainment and both BMI and self-rated health, and weak evidence for genetic confounding of the relationship between college degree attainment and depression.

LITERATURE REVIEW

The Relationship between Educational Attainment and Health

Sociologists and other social scientists have long been interested in documenting the associations between social conditions and various health outcomes. These associations persist across time and space and are some of the most robust findings in the social sciences. Given the strong associations between social conditions and health outcomes, social conditions such as education have been conceptualized as a fundamental cause of health inequality by social scientists. Link and Phelan's Fundamental Cause Theory states that social conditions can provide individuals with resources that allow them to both minimize disease incidence and to maximize positive health outcomes (Phelan et al. 2010). This process can occur through multiple mechanisms which is why the association tends to remain across many different contexts (Link and Phelan 1995). Link and Phelan define resources broadly as "money, knowledge, power, prestige, and the kinds of interpersonal resources embodied in the concepts of social support and social network" (1995:87). While the types of resources or characteristics that Link and Phelan identify certainly are affected by social conditions such as education, many of them plausibly have genetic sources as well.

In order to provide stronger evidence for the argument that education is a social determinant of health, researchers have moved beyond demonstrating associations between education and various health outcomes to better understanding the causal relationship. These efforts have included controlling for possible selection or confounding factors, using within identical twin pair fixed-effect methods, or designing studies using natural experiments or

instrumental variable models that allow for the estimation of a causal effect of education (Amin, Behrman, and Kohler 2015; Boardman and Fletcher 2015; Davies et al. 2018; Fletcher 2015; Kane et al. 2018; Zheng 2017).

Studies using monozygotic or identical twins rely on the fact that monozygotic twins are genetically identical, and they assume that their family backgrounds will be equal if they grow up in the same home. Past studies using within-twin fixed effects designs have yielded mixed evidence regarding the existence of a causal effect of education on health (Amin et al. 2015; Behrman et al. 2011; Behrman, Xiong, and Zhang 2015). Amin and colleagues used this approach to estimate the causal effect of education on a number of health outcomes in three separate samples of identical twins in the United States, and found that education did not have a significant effect on any of the health outcomes, with the exception of self-rated health (2015). This lends support to earlier findings from a Danish twin cohort where no association between education and mortality was found after controlling for unobserved factors using the same within-twin pair estimator (Behrman et al. 2011). However, in a Chinese twin cohort, Behrman and colleagues found that negative associations between education and smoking behaviors remained after controlling for unobserved factors common to a twin pair. They also found that controlling for unobserved factors uncovered a positive causal effect of education on mental health that did not appear in cross-sectional associations (Behrman et al. 2015). As this was the first use of the within-twin fixed effects methodology in a developing country context, it suggests that patterns found in the United States or other developed countries may not apply universally.

While the use of identical twins to estimate the effect of education on health is useful in that it controls for unobserved genetic factors that are the focus of this study, the use of identical

twins for causal inference is not without its limitations. The external validity of studies using samples of twins is suspect, as twins are a selective group who are not representative of the population at large (Boardman and Fletcher 2015).

Another approach to estimating the causal effect of education on health is the use of instrumental variables or regression discontinuity designs. Most of the studies that use exogenous variation in education to estimate a causal effect rely on the increasing age of mandated schooling that unfolded as the 20th century progressed (Fletcher 2015). Results from studies using these approaches in the United States generally find support for a causal relationship between education and health in the United States. In perhaps the most well-known example of this, Lleras-Muney (2005) estimates a causal effect of education on mortality in the expected direction. While these findings have been questioned by Mazumder (2010), more recent evidence suggests that the original findings hold for several measures of health (Fletcher 2015). In an even more recent study using UK Biobank data, Davies, Dickson, Smith, van den Berg, and Windmeijer (2018) find that additional education has an effect on both the risk of diabetes and mortality.

However, studies using the raising of the minimum school leaving age also suffer from various limitations. Such a strategy results in the estimation of the effect of an additional year of high school education, in most cases, on health and mortality in later life. However, causal effects from receiving a high school diploma or from additional years of college education cannot be estimated using such approaches, and past research suggests that these effects may differ from those of additional years of high school education (Montez, Hummer, and Hayward 2012). For example, Montez and colleagues find that the association between educational attainment and mortality is linear until high school graduation, which results in a step-change in

mortality risk. Additional years of education beyond high school are again linearly associated with mortality, but the effect of each additional year is larger (Montez et al. 2012).

While a popular identification strategy, the raising of the minimum school leaving age is not the only instrument that can identify the effect of education or other measures of human capital. As an example, Kane et al. (2018) find causal effects of human capital on health, and metabolic syndrome more specifically, by using a number of measures of neighborhood quality as instruments. This instrumental variable application, and others like it, do not suffer from the criticism above, in that the exogenous variation does not only come from an additional year of high school education. However, instrumental variable methods regardless of the instrument only estimate the local average treatment effect (Deaton 2009). In other words, we can only know the average effect of education for individuals whose education is affected by the instrument.

Both approaches discussed above are powerful and have much to offer for our understanding of the causal effect of education on health. However, as discussed, they are not without their limitations and study of the education-health relationship should not be limited to those approaches. In fact, researchers regularly employ methods of covariate adjustment to attempt to control for possible confounding factors. Many of the recent studies concerned with the causal relationship between education and health that control for possible selection or confounding factors use propensity score methods (Rosenbaum and Rubin 1983). Two recent studies use this approach to attempt to account for selection into a college degree and to estimate the causal effect of a college degree on obesity (Lawrence 2017), and self-rated health and depression (Zheng 2017). Importantly, neither study includes any genetic measures when modeling selection into receipt of a college degree. Because they each aim to characterize possible confounders in the relationship between a college degree and the health outcomes I

study here, they can be informative regarding important control variables to include in my analysis. This is especially true for Lawrence as she uses Add Health data and includes 54 measures at Wave I (2017). In order to account for as many selection factors as possible and because it is the basis of propensity score analyses, both Zheng and Lawrence include a "kitchen sink" of potentially relevant variables measured before the respondents enter college (Lawrence 2017; Zheng 2017).

After accounting for selection into college degree attainment, they find that the estimated effect of a college degree on obesity is reduced by 54% (Lawrence 2017), while the estimated effects of a college degree on self-rated health and depression are reduced by about 53% and 70% respectively (Zheng 2017). These results make it clear that non-genetic measures substantially confound the relationship between education and health. However, many of the variables included in these analyses are likely to be partial proxies for direct genetic measures, as both authors mention that genetic factors could be possible confounders. For example, Zheng (2017) includes measures of both cognitive and non-cognitive skills, and earlier measures of health. Lawrence (2017) also includes earlier measures of health and a measure of cognitive ability, as well as a scale of high school grades.

Recently, researchers have begun to employ PGSs to test for genetic confounding in a number of contexts (Gaydosh et al. 2018; Liu 2019; Wertz et al. 2019). Below, I explain why and when genetics could be an important confounder of the relationship between education and health.

Genetic Correlation (rG)

Education and the health outcomes such as BMI, self-rated health, and depression, have been shown to be at least partially heritable, based on findings from research using both twin

models and molecular genetic methods (Boardman et al. 2015; Haberstick et al. 2010; Johnson et al. 2002; Leinonen et al. 2005). There is also evidence that education and each of the above health outcomes have at least some common genetic influences, as indicated by genetic correlation (rG). Genetic correlation is a measure of the average correlation of the effect sizes of individual genetic variants across two different traits (Bulik-Sullivan, Finucane, et al. 2015). Wedow, Zacher, Huibregtse, Harris, Domingue, and Boardman (2018) provide an excellent explanation of rG by example which I borrow from in my explanation below. In order to understand rG intuitively, it may be helpful to think of situations in which two phenotypes are almost perfectly genetically correlated or are genetically independent from one another. BMI and waist circumference, are two anthropometric measures that are usually highly correlated, and also share an estimated rG of over .9 (Zheng et al. 2017). This means that almost all the genetic variants influencing BMI also influence waist circumference in the same direction. We can also consider schizophrenia and smoking behavior, which are associated but have an rG that cannot be distinguished from zero (Bulik-Sullivan, Loh, et al. 2015). This means that any association between these traits is not due to genetic factors.

In an early attempt to use molecular genetic data to explore the possibility for genetic confounding in the education-health association, Boardman et al. (2015) tested for the possibility of genetic confounding in the three education-health relationships by estimating rG between education and each measure of health in the Health and Retirement Study (ages 50 and above) genetic data. Using genome-wide complex trait analysis (GCTA), a method for estimating rG using molecular genetic data in unrelated individuals, they estimate a positive rG between education and self-rated health and a negative rG between education and depression, but find no evidence for rG between education and BMI (Boardman et al. 2015).

The disadvantage to estimating rG using GCTA is that it requires very large sample sizes to produce precise estimates and requires individual genetic data. For example, with a sample size of 4,233, Boardman and colleagues (2015) estimate a genetic correlation between education and depression of -0.746, but the 95% confidence interval of their estimate ranges from -1 to -0.201. The recent development of cross-trait linkage disequilibrium (LD) score regression has made rG estimation possible using summary statistics from genome-wide association studies (GWAS) rather than individual genetic data (Bulik-Sullivan, Finucane, et al. 2015). This allows for more precision in the estimation of rG because researchers can exploit the summary statistics from very large GWAS. Using this method, researchers have since estimated an inverse rG between both education and BMI (Okbay et al. 2016) and education and depression (Wray et al. 2018), as well as a positive rG between education and self-rated health (Harris et al. 2017). The negative rG between education and BMI indicates that genetic variants that have a positive effect on education tend to have a negative effect on BMI on average. However, estimates of rG do not in themselves provide evidence for genetic confounding as they can reflect either mediated or biological pleiotropy as discussed below.

Pleiotropy

These estimates of rG between education and each health outcome provide suggestive evidence for the presence of pleiotropy, which occurs when one or more genetic variants influence two or more separate traits or phenotypes (Solovieff et al. 2013). Both Boardman et al. (2015) and Wedow et al. (2018) also outline a clear distinction between two different types of pleiotropy, which I adopt here because of its utility for explaining when rG could result in genetic confounding. Figure 1 can be referenced for a visual representation of the concepts discussed. One possibility is that the genetic variants associated with both education and the

health outcome have an indirect effect on the health outcome that is mediated through education. This is defined as mediated pleiotropy (Figure 1B). If measured education completely mediates the relationship between the pleiotropic variants and the health outcome of interest, rG between education and that health outcome would not contribute to genetic confounding of the education-health relationship.

The other possibility is that the genetic variants associated with both education and the health outcome influence a trait that is a proximate determinant of both education and the health outcome, such as conscientiousness. That trait would then have a direct effect on both education and the health outcome. This is defined as biological pleiotropy (Figure 1A). In the case of biological pleiotropy, because the effect of the genetic variants is not mediated through education, rG is likely to confound the education-health association. A potential partial solution, which I test in this paper, is to include polygenic scores for education and the health outcome as control variables in the model.

Polygenic Scores (PGSs)

An ideal measure to test for genetic confounding would capture all common genetic influences of education and each health outcome. Because there are no available methods that can do this, I rely on three separate polygenic scores (PGS) for educational attainment, BMI, and dep[r]ession. A PGS is an additive measure of the effects of individual variants across the genome on a phenotype. It can be conceptualized as an additive whole-genome measure of genetic influence on some outcome (Belsky and Israel 2014; Dudbridge 2013). I include a PGS for education in all models. I also include PGSs for BMI and depression in models that include BMI and depression as dependent variables to ensure that as much of the association between common genetic effects is controlled for as possible. Because a similarly high-powered GWAS

for self-rated health is not available, I include both the PGS for BMI and for depression in all models that include self-rated health as a dependent variable. This is motivated by the understanding that self-rated health tends to reflect both physical and mental health (Bailis, Segall, and Chipperfield 2003).

PGSs are also able to provide some information about the type of pleiotropy that is occurring. They cannot provide any information about pleiotropy at the level of an individual genetic variant, but if a PGS is associated with both college degree attainment and health, the reason for that association is informative for pleiotropy at the level of genome-wide genetic influence. If the PGS has an indirect effect on health mediated by college degree attainment, this is evidence for mediated pleiotropy. If the PGS has direct effects on both college degree attainment and health, this is evidence for biological pleiotropy as previously defined.

While controlling for PGSs does not result in a causal estimate of the effect of a college degree on health, we still learn a number of things about how genetics factor into the relationship between college degree attainment and health. If we find evidence for genetic confounding, this tells us that the observed rG between education and health is at least partly driven by biological pleiotropy, and we can estimate the proportion of the effect that is confounded. This will be informative for researchers studying the education-health relationship in the future when making methodological decisions to account for potential confounders. If we do not find evidence for genetic confounding, this tells us that the observed rG is driven mostly by mediated pleiotropy, and that concern about confounding from rG estimates alone may not always be warranted.

DATA AND MEASURES

I draw on data collected by the National Longitudinal Study of Adolescent to Adult Health, hereafter referred to as Add Health (Harris et al. 2009). Add Health is a nationally representative longitudinal study of adolescents in the United States who were in grades 7-12 in 1994-1995 during Wave I. Data are available for four waves, and data for Wave V of the study was recently released. Because the Wave V data was released during my analysis, I use data from Waves I, II and IV. In Wave IV, respondents were between the ages of 24 and 32. Genotyping was performed in Wave IV, and of the 15,701 participants in Wave IV, approximately 12,200 were genotyped using two Illumina platforms. Approximately 80% of the sample were genotyped using the Illumina Omni1-Quad BeadChip and 20% were genotyped with the Illumina Omni2.5 - Quad BeadChip. Genotyped data were available on 609,130 SNPs for 9,974 individuals after applying standard quality control procedures. Quality Control Analysis of Add Health GWAS Data documentation (Highland et al. 2018) provides more information on the genotyping procedures in Add Health. To account for population stratification, I restrict my analysis to individuals of European genetic ancestry, bringing the size of the analytic sample to 5,629 after excluding individuals without valid sampling weights (Braudt and Harris 2018).

Add Health is an ideal data set to use both because of the available genetic measures in the data set and because the sample consists of a younger cohort that was representative of the United States middle- and high-school enrolled population at the time of sampling in 1994-95. Add Health provides rich environmental and health measures over the first part of the life course,

while also providing genetic data on a large portion of its respondents. In addition, in some respects it has an advantage over other larger datasets with genetic data like the UK Biobank in that it began with a nationally representative sample of adolescents. While there are other social science datasets with genetic data that are nationally representative like the Health and Retirement Study, Add Health provides the opportunity to examine a cohort who has more recently experienced their educational attainment. Given evidence that genetic effects are conditional on the environmental context (Tropf et al. 2017), it is important to assess genetic confounding for different cohorts who experience their educational environments in different time periods. It is important to note that while Add Health was designed to be nationally representative, my analysis is restricted to individuals with European genetic ancestry. Therefore, the results cannot be generalized to populations with non-European genetic ancestry.

I measure educational attainment in Wave IV in order to capture the highest level of education that respondents have attained at the age of 24-32. I use receipt of college degree rather than years of education in order to facilitate the propensity score analysis described below. While neither parameterization is likely to perfectly represent the functional form of the relationship between education and health, there is some evidence that college education is uniquely important for health (Montez et al. 2012).

Because Wave V data on the full sample were not available at the time of this analysis, I also measure the three health outcomes of interest at Wave IV. While it would be ideal for the measurement of education to occur at a time point before the health measurements, most of the respondents will not have finished their educations at Wave III. I use the constructed measure of body mass index (BMI) as a continuous variable. This was calculated using the height and weight of respondents measured in Wave IV. Anthropometric measures like height and weight,

and BMI calculated from those measures, have been shown to be highly reliable in Add Health (Hussey et al. 2015). To measure self-rated health, respondents were asked “In general, how is your health?” and were able to choose between “excellent”, “very good”, “good”, “fair”, and “poor”. I treat self-rated health as an ordinal variable.

To measure depression, I use a measurement model with four indicators drawn from the Center for Epidemiological Studies Depression Scale (CES-D Scale) (Radloff 1977). The subset of items I am using were originally part of a measurement model proposed and validated by Perreira et al. which, in line with measurement theory, is composed of only the effect indicators in the original scale while excluding any causal indicators or outcomes of depression (Perreira et al. 2005). Because this is a rare case in which a measurement model has already been tested and validated in the dataset I am using, it makes sense to begin with this as my initial specification. Figure 2 is the path diagram representing the model. Respondents were asked, “How often was the following true during the past seven days?” and then given the following four prompts. “You could not shake off the blues, even with help from your family and your friends.” “You felt depressed.” “You felt happy.” “You felt sad.” Respondents were given a “0” if they chose “never or rarely”, a “1” if they chose “sometimes” a “2” if they chose “a lot of the time”, and a “3” if they chose “most of the time or all of the time”. I treat each of these indicators as an ordered categorical variable.

Because genotype is determined at birth and PGSs are a summary genetic measure, it may seem unnecessary to include control variables if we are interested in estimating effects of the PGSs. However, the effects of PGSs are known to be confounded by family background because parents both transmit their genetics to their children and influence their educational attainment and later health (Young et al. 2019). Without including controls for family

background, the effect of genetic confounding could be overestimated. Furthermore, a secondary aim of this study is to investigate genetic confounding that would not be captured by variables that researchers commonly have access to in large social science surveys.

To accomplish both goals, I estimate two sets of models. The first set of models includes covariates only meant to control for the confounding effects of family background at Wave I. I do not include individual characteristics of the adolescents in this set of models because they may mediate the effects of the PGSs. In the second set of models I attempt to control for many potential confounders of the relationship between college degree attainment and health in the Wave I and II data.[1] Because Lawrence (2017) previously attempted a similar strategy to account for selection into college degree attainment in Add Health, the first set of models includes the following subset of the variables used by Lawrence.

Parents' education is measured using an average of resident parent years of education. When data do not exist for both parents, I use the education level of one parent. Parent occupation is measured using two binary indicators of whether the resident mother and father each have professional jobs. As a measure of family income, I follow Lawrence (2017) and use a categorical income-to-needs ratio variable. This ratio is calculated by dividing the household income by the poverty threshold in 1994 for the adolescent's household size. I create a separate missing category for this variable because 18% of the respondents have missing information. Other household level controls at Wave 1 include whether the respondent had health insurance, whether the parent interviewed received public assistance, whether there was a smoker in the household, and the household size. Controls for parent health behaviors include whether the

[1]While many of these covariates likely confound the relationship between college degree attainment and health, they also likely mediate the effect of the PGSs on college degree attainment and health respectively. I therefore do not include all covariates used by Lawrence (2017) in the first set of models because it would result in the underestimation of genetic confounding.

parent smoked and the frequency of parent heavy drinking. Age, sex, and whether the respondent was born in the United States are also included as control variables in all models.

As a control variable for the second set of models, I estimated propensity scores using the list of variables in Table 1 under the "variables for propensity scores" heading, which includes the variables described above. These are the same variables included in Lawrence's (2017) analysis. To handle missing data, ten datasets were imputed using the data imputation command in Mplus version 8.2.

To measure genetic contributions to the outcomes of interest in this study, I rely on three publicly available PGSs from Add Health for educational attainment, BMI, and major depressive disorder, respectively (Braudt and Harris 2018; Lee et al. 2018; Locke et al. 2015; Wray et al. 2018). Each polygenic score represents the cumulative additive genetic influence on the phenotype of interest (Belsky and Israel 2014). The PGS for an individual i is calculated as:

$$PGS_i = \sum_{j=1}^{k} \beta_j SNP_{ij}$$

where, SNP_{ij} is the number of alleles of the j^{th} SNP for the i^{th} individual and β_j is the estimated association between SNP j and the phenotype as reported in the summary statistics for a GWAS of that phenotype based on an independent sample. The PGSs are then standardized to have a mean of 0 and a standard deviation of 1. For more information on the calculation of the PGSs used in this study, see Braudt and Harris (Braudt and Harris 2018).

Even in samples restricted on ancestry, population stratification can still create spurious associations between PGSs and outcomes (Price et al. 2006). This occurs because ancestral differences within populations may be associated with but causally unrelated to outcomes of interest, and a failure to control for these differences can result in biased estimates of the effects

of PGSs. A common method for addressing this, which I use here, is to regress the original PGSs scores on the first ten ancestry-specific principal components of the genetic data, save the residuals, and then standardize them (Belsky et al. 2018; Selzam et al. 2019).

METHODS

While past studies have estimated non-zero rG between education and these health outcomes in other datasets, it is preferable to provide evidence of this in Add Health before moving on to more complicated analyses. I do this by estimating bivariate correlations between all variables of primary interest in the analysis, which include the PGSs, college degree attainment, and the three health outcomes.[2]

I begin the analysis examining the relationship between receipt of a college degree and depression with the estimation of the confirmatory factor analysis model (measurement model) for depression described above and illustrated in Figure 2. "You felt depressed" is used as the scaling indicator.

I then estimate a series of structural equation models (SEMs). The general path diagrams representing these models are in Figure 3. The purpose of each is to test whether genetics confound the relationship between receipt of a college degree and each of the three health outcomes (depression, BMI[3], and self-rated health). The right side of each diagram includes variables measured in Wave 4 and depicts the relationship between education and the health outcome of interest. There are a few reasons why SEM is preferred for this analysis. For one, depression is best modeled as a latent variable to control for measurement error, which is easily done in an SEM framework. Because the depression models are overidentified, I am also able to take advantage of various model fit statistics. Furthermore, the tradition of presenting path

[2]I declare college degree attainment and self-rated health to be categorical variables before estimating the correlations.
[3]BMI models exclude individuals who stated they were or might be pregnant during Wave IV.

diagrams with SEM analyses, which I do here, helps with interpretation of the model results. The models that do not include the latent measure of depression are fully recursive, so the results would be no different if they were estimated one equation at a time using standard regression models. Finally, as explained below, I exploit a feature of structural equation modeling software used for mediation analysis to formally test for a confounding effect.

In the first round of models, the controls on the left sides of the diagrams are possible confounders of the effects of the PGS and the reason for including them is discussed above. All control variables are entered separately in the models and can correlate but are represented in the path diagrams as one observed variable for parsimony.

In the second round of models, I replace the smaller set of control variables with a propensity score to represent the joint confounding effects of all covariates that were included in the first round of models plus the variables under the heading "variables for propensity scores" in Table 1. Lawrence (2017) applies propensity scores estimated using these same covariates to account for selection into a college degree, and she acknowledges in her limitations section that she was not able to account for genetic factors. Some of the variables used to estimate the propensity scores will likely mediate the influence of the PGSs in this analysis. For example, Wave I measures of the health outcomes are included in the estimation of the propensity scores. However, the purpose of the propensity score analysis is to see whether any evidence of genetic confounding remains after controlling for a wide range of possible confounders measured at Waves I and II. This is therefore a conservative test of genetic confounding, as many of the covariates included in the propensity score model will inadvertently capture some genetic influence. The propensity scores used in the analysis were generated by estimating a probit

regression model of the receipt of a college degree on relevant variables listed in Table 1 and saving the propensity scores. This was done separately for each of the 10 imputed datasets.

The direct effects of the education PGS on the health outcome and the direct effect of the relevant health PGS on receipt of a college degree are the primary coefficients of interest for this analysis. For example, there is likely to be a significant non-zero direct effect of the education polygenic score on receipt of a college degree. This would not indicate genetic confounding. If this same polygenic score also influenced depression however, this would then make education linked genetics a possible confounder in the relationship between receipt of a college degree and depression. We would also expect the PGS for BMI to have a significant non-zero direct effect on BMI, but if it has a significant non-zero direct effect on receipt of a college degree, this indicates possible genetic confounding.

For each health outcome under consideration, I fit models using two different specifications. Path diagrams for each specification are in Figure 3 where they are labeled as Model 1 and Model 2. The equation predicting the health outcome in Model 1 can be understood as a regression of the health outcome on receipt of a college degree with controls for observed covariates, no PGSs are included. The effect of a college degree on health in this model will be treated as a baseline for comparison with the effects from the second model.

Model 2 adds direct paths from both PGSs to both education and the health outcome, which is the test of genetic confounding. If one PGS has a direct effect on both education and the health outcome, it confounds the association. This would provide evidence for genetic confounding. Model 3 includes PGSs for both education and the health outcome but assumes that each polygenic score only has a direct effect on the outcome it is optimized to predict. For example, Model 3 predicting depression includes a direct path from the PGS for depression to

depression, but not to education. Likewise, it includes a direct path from the PGS for education to college degree, but not to depression.

To formally test for genetic confounding, I exploit a common feature of structural equation modeling programs that is commonly used for mediation analysis. To test for mediation in structural equation models, it is common to estimate an indirect effect by taking the product of the direct effects that lie along the path of meditation (Bollen 1989). Variability of the indirect effect can then be estimated using the delta method or bootstrap methods (Bollen and Stine 1990; Sobel 1982). It is well known that mediation and confounding are statistically equivalent, and that the only way to distinguish between the two processes is by using theoretical knowledge about the causal ordering of variables (MacKinnon, Krull, and Lockwood 2000). Therefore, I test for the significance of the confounding effect by testing for the significance of an indirect effect of college degree attainment on health in an "incorrect" model where the PGSs serve as mediators rather than confounders. Each equation in this model includes as covariates those control variables that are present in the corresponding "correct" model.

For symmetry between the analyses using observed covariates and propensity scores, I use the propensity score as a covariate to adjust the models described above. While this is a common approach in the literature, some have discouraged this use of the propensity score because it assumes a linear relationship between the propensity score, the treatment, and the outcome (Austin 2011). As a robustness check, I use the propensity score to calculate inverse probability of treatment weights (IPTW) to weight the sample so that it is balanced on the hypothesized confounders (Austin 2011). I then use IPTW regression models to estimate the

effect of a college degree on the health outcomes.[4] I can then test for genetic confounding in cases where it remains after including the propensity score as a covariate in the model.

Estimation

All analyses were performed in Mplus version 8.2. For all depression and BMI models, I first estimated them using a robust weighted least squares estimator using a diagonal weight matrix (WLSMV in Mplus) to handle both the categorical nature of the depression indicators and the college degree variable and because this estimation method provides a wide array of fit statistics. However, when using WLSMV in models with categorical mediators, the continuous latent response variable rather than the categorical itself is used as the covariate in the regression of the outcome on the mediator. Because college degree is a categorical mediator in all models, this would affect the interpretability of the college degree coefficient. In models estimated using WLSMV, the college degree coefficient gives the expected change in the outcome that corresponds to a one-unit change in a latent variable underlying the binary college degree variable. This coefficient is not directly interpretable as the effect of a college degree. To solve this problem, I also estimated all depression and BMI models using maximum likelihood with robust standard errors (MLR in Mplus), which uses the observed mediator as the covariate and allows for the direct interpretation of the effect of a college degree.

For models predicting self-rated health, I estimate each equation of the model separately which is reasonable because the model is fully recursive. I use MLR to estimate the equation predicting receipt of a college degree as was done in other models, but I use WLSMV to estimate the equation predicting self-rated health. This is done because self-rated health is an ordinal

[4]Individuals with propensity scores equal to zero or one were excluded from these analyses. Individuals with propensity scores very close to zero or one can be assigned IPT weights that are unreasonably large. For this reason, I follow Lawrence (2017) and top code the weights at 14.

variable and WLSMV in Mplus provides y-standardized coefficients. Y-standardizing the coefficients allows for an interpretation of coefficients that is analogous to linear regression. It also allows for the comparison of the effect of a college degree across different self-rated health models, as this method of interpretation is not affected by the rescaling problem that comes with comparing odds ratios across different nonlinear probability models (Breen, Karlson, and Holm 2018). However, I cannot estimate the full model using WLSMV because of the problems discussed above.[5] All analyses account for the complex sampling design of Add Health through sampling weights, stratification, and clustering.

[5]I used a binary measure of education to facilitate the creation of propensity scores, but it is important to note that does not likely capture the entire effect of education. Because of this, it is possible that the underlying continuous measure could confound the college degree effect.

RESULTS

Beginning with the indirect tests of rG, we can see in Table 2 that the educational attainment PGS correlates with the BMI PGS at -0.182 and the depression PGS at -0.140. These are relatively weak in magnitude but statistically significant. This suggests that on average, the genetic variants associated positively with educational attainment tend to be negatively associated with BMI and depression, at least for the variants captured in the PGS. On the other hand, the correlation between the PGS for BMI and the PGS for depression is -0.033 and not statistically significant, suggesting that there is no clear relationship between the genetic variants associated with BMI and depression. While there may be relationships between individual variants captured by the PGSs, a correlation between the two PGSs would not necessarily show this.

We can also examine the correlations between the PGSs and relevant outcomes in the same table. Starting with the relationships between the health PGSs and college degree attainment, the BMI PGS and college degree attainment are significantly correlated at -0.129 and the depression PGS and college degree attainment are significantly corelated -0.122. The education PGS is also significantly correlated at -0.072 with depression, -0.063 with BMI, and 0.117 with self-rated health. Taken together with the past studies of rG previously reviewed, these results provide evidence for rG between educational attainment and these three health outcomes in the Add Health cohort. While the correlations estimated here are modest, they are an indirect indication of rG and motivate investigation of potential genetic confounding.

The first model to be fit is a confirmatory factor analysis of depression as outlined above. As shown in Table 3, the model fits the data very well. The TLI and CFI are both equal to 1, and the RMSEA is 0. The chi-squared statistic[6] is small and the BIC (Bollen et al. 2014; Raftery 1995) is around -17. The R^2 values of the indicators range from .519 to .879, which means that the latent variable of depression is explaining a relatively large amount of the variance in the underlying indicators. After determining that the measurement model fits well, we can now turn to the fit of the structural equation models predicting depression.

The main paths of interest in this model are the direct paths from the polygenic scores to education and depression, so each specification of the model sets some of these paths to be estimated and some others to be zero. The columns in Table 3 are labeled according to the model specifications pictured in Figure 3. Model 1, which constrains the direct effects of both PGSs on both college degree and depression to be zero, has the worst fit according to all the fit statistics presented. Model 2, which frees all paths that were constrained in Model 1, has a fit that is almost indistinguishable from Model 3, which constrains the effect of the depression PGS on college degree and the effect of the education PGS on depression to zero. This means that the structure of Model 3 is assuming that there is no direct effect of the polygenic score for education on depression, and there is no direct effect of the polygenic score for depression on education. Model 2 however, is assuming that the PGSs operate as confounders, having direct effects on both college degree attainment and depression. Because the fit statistics between Model 2 and Model 3 are very similar, examining the regression coefficients of the models

[6]All fit statistics reported for the depression models are means averaged over the ten imputed datasets. For this reason, Mplus does not provide a significance test for the chi-squared statistic and a formal statistical comparison of model fit cannot be done. (Muthen and Muthen n.d.)

provides more information to judge whether genetic factors are a possible confounder of the relationship.

The results for the models predicting depression are presented in Table 4. In Model 1, which does not include PGSs, having a college degree is associated with a 1.352 unit decrease in depression. In Model 2, both PGSs are included as potential confounders of the relationship, as a confounder should have a significant direct effect on both education and depression. While the education PGS only has a direct effect on college degree attainment, the depression PGS has a significant direct effect on both college degree attainment and depression, providing some evidence for genetic confounding by the depression PGS. If the PGSs were confounders of the association, a failure to control for them would result in a biased estimate of the effect of education and we would therefore expect the effect estimate to change as they were added to the model. In Model 2, the college degree coefficient is associated with a 1.301 unit decrease in depression, which is a change of about 4% from model 1. Using the method described above to estimate the total indirect effect, the estimate of the confounded effect was not determined to be statistically significant ($p = 0.126$). However, a small specific indirect effect through the depression PGS was detected ($\beta = -0.035$; $p = 0.046$), which suggests that the depression PGS could operate as a partial confounder of the relationship between college degree attainment and depression. Overall, the evidence for genetic confounding here is relatively weak, as is the evidence for biological pleiotropy as an explanation for rG. However, failing to control for PGSs in this context could result in a slight overestimate of the effect of a college degree on depression. In the relationship between college degree attainment and depression, observed college degree attainment also mediates the relationship between the education PGS and depression, which is evidence of mediated pleiotropy.

For the models focused on the relationship between college degree attainment and BMI and between college degree attainment and self-rated health, Model 2 is exactly identified and so it has no fit statistics. However, it is the saturated model that is automatically compared with Models 1 and 3 through their fit statistics. The fit statistics for the BMI models are in Table 5. Model 1, which constrains all the effects of the PGSs, has a terrible fit. Model 3, which freely estimates the effects of the PGSs only on their corresponding phenotypes, has a much better fit, but still has room for improvement, especially on the TLI.

The regression results from the models predicting BMI are displayed in Table 6. After controlling for the education PGS and the other observed covariates in Model 2, the BMI PGS has a significant direct effect on both college degree attainment and BMI. Furthermore, we can compare the regression coefficients on college degree between Model 1 and Model 2. In Model 1 graduation from college is associated with a decrease in BMI of 1.652 units, while in Model 2 it is associated with a decrease in BMI of 1.425 units. The absolute value of the coefficient decreases by about 14% ($p = 0.008$). This provides evidence for genetic confounding, as the estimated effect of education weakens significantly after controlling for the PGS. The substantial statistically significant decrease in the effect of a college degree on BMI between the two models, the significant direct effect of the BMI PGS on college degree attainment and BMI in Model 2, and the less than ideal fit of Models 1 and 3 provide evidence for genetic confounding of the relationship between college degree attainment and BMI. While we find evidence for genetic confounding, and therefore biological pleiotropy likely partially contributes to the rG between education and BMI, we can also see that the PGS for education has an indirect effect on BMI mediated through observed education, which is evidence for mediated pleiotropy.

Turning to the fit statistics for the self-rated health models in Table 7, we can observe that they are patterned similarly to the fit statistics for the BMI models. Model 1 again has a terrible fit, and Model 3 is better but shows room for improvement. This provides suggestive evidence for genetic confounding, but the regression estimates provide additional evidence. In Model 1 predicting self-rated health (Table 8), graduating from college is associated with a 0.449 ($0.472/\sigma_{y^*}$) standard deviation increase in self-rated health. In Model 2, after controlling for the education, BMI, and depression PGSs, graduating from college is associated with a 0.421 ($0.445/\sigma_{y^*}$) standard deviation increase in self-rated health.

Given that the model predicting self-rated health is an ordinal probit, we cannot directly calculate the percent decrease in the effect of education on self-rated health from the coefficients as we move from Model 1 to Model 2 in Table 6. This approach is inappropriate in nonlinear probability models such as ordinal logistic regression because adding additional variables to these models will affect both the residual variance of the model and the true effects of the variables in the model. The change in the coefficient captures both effects, so it does not allow us to distinguish between true confounding and a change in the residual variance. This is not a problem in linear regression because the residual variance is estimated separately (Breen et al. 2018). In order to retrieve unbiased estimates of genetic confounding, I divide the college degree coefficient by the estimated standard deviation of the continuous latent variable underlying self-rated health (Breen et al. 2018).

Using the product coefficient method described above, about 6% ($p < 0.001$) of the effect of college degree attainment on self-rated health is confounded by PGSs for education, BMI, and depression. In Model 2, the PGS for BMI also has a statistically significant direct effect on both college degree attainment and self-rated health. While the percent confounded in this relationship

is smaller than in the college degree and BMI relationship, failing to control for common genetic factors results in a small overestimate of the effect of a college degree on self-rated health. As in the models predicting depression and BMI, we also observe evidence for mediated pleiotropy in the relationship between college degree attainment and self-rated health, as the education PGS has an indirect effect on self-rated health that is mediated through observed education.

The above analyses are informative because they only include additional controls that represent the respondent's family background in Wave I. They are unlikely to mediate the effects of the PGSs in this case. However, given the rich information available in Add Health and in many other social science surveys, one might wonder whether the evidence for genetic confounding presented above remains after controlling for variables that may inadvertently capture genetic effects. This question motivated the analyses using propensity scores reviewed below.

Table 3 also includes the fit statistics from the three depression models that include a propensity score. The only difference between these models and the previous three depression models is that a propensity score, saved from a probit regression of college degree attainment on the covariates in Table 1, replaces the covariates in the previous models.[7] Unlike the first round of depression models, Model 1, which constrains the effects of the PGSs to zero, is no longer clearly the worst fitting model, and all three of the models seem to fit well. However, if one model were to be chosen based on the fit statistics alone, Model 3 seems to be the best fitting model. As a reminder, Model 2 allowed all coefficients to be freely estimated, while Model 3 constrained the effect of the education PGS on depression and the effect of the depression PGS

[7]The covariates from the first round of models are included as part of this list.

on college degree attainment to be zero. This provides some evidence that the small confounding effect of the depression PGS might have been blocked by the propensity score.

We can also examine the regression coefficients for these models in Table 9. The first thing to notice is that the depression PGS no longer has a significant direct effect on college degree attainment or depression in Model 2. In addition, graduating from college is associated with a .632 unit decrease in depression in Model 1, which is a 53% decrease from the original specification of Model 1 that did not include a propensity score. This is a much larger decrease than occurred after including the PGSs in the original models, suggesting that the propensity score is accounting for more confounding than were the PGSs. After controlling for the PGSs in the propensity score models, the college degree coefficient still declines by about 2% from Model 1 to Model 2, but neither the total indirect effect nor either of the specific indirect effects estimated using the product method are statistically significant. When we also consider the lack of differences in fit across models and the non-significant direct effect of the depression PGS on college degree attainment and depression in Model 2, we can conclude that any small confounding effect of the depression PGS was blocked by the propensity score.

For the propensity score models predicting BMI, Model 1 has a noticeably worse fit than Model 3, but Model 3 fits well and only has a small amount of room for improvement (Table 5). This is suggestive evidence that the propensity score may be capturing at least some of the genetic confounding effect from the first series of BMI models. The estimated effect of a college degree is also much smaller in Model 1 compared to the same model without the propensity score. In Table 10, graduating from college is associated with a 1.071 unit decrease in BMI, which is a 35% absolute decrease from the same coefficient reported in Table 6. As in the depression models, the propensity score seems to be capturing a much larger confounding effect

than the PGSs alone. In Model 2, the effect of the BMI PGS on college degree attainment is no longer significant like it was in Model 2 without the propensity score and graduating from college is associated with a 0.964 unit decrease in BMI. Moving from Model 1 to Model 2 in Table 8, the coefficient on college degree decreases by about 10%. While this change might seem relatively large, the total indirect effect estimated through the product method is not statistically significant (p = 0.217). As with depression, it appears as if controlling for the propensity score blocks the genetic confounding effect that was found in the previous round of models.

The fit of the models predicting self-rated health with propensity scores is like that of the BMI models with propensity scores, in that Model 3 fits very well, while Model 1 does not fit as well (Table 7). Again, this suggests that the propensity score may be blocking at least a portion of the previously identified genetic confounding. In Table 11, we can see that the y-standardized effect of a college degree on self-rated health decreased from 0.449 ($0.472/\sigma_{y^*}$) to 0.285 ($0.300/\sigma_{y^*}$), a decrease of 37%, when comparing coefficients from Model 1 before and after adding the propensity score. The decrease after adding the propensity score is like that observed for depression and BMI. We also notice that the BMI and depression PGSs are no longer significant predictors of college degree attainment in Model 2. Moving from Model 1 to Model 2 however, the y-standardized effect of a college degree on self-rated health decreases from 0.285 ($0.300/\sigma_{y^*}$) to 0.274 ($0.290/\sigma_{y^*}$), a statistically significant decrease of about 4% (p = 0.031). While the propensity score does block some of the genetic confounding, a statistically significant but small confounding effect due to genetic factors remains.

The models discussed above that included the propensity score as a covariate assume that the model is specified correctly, which in this case means that the propensity score is linearly

related to college degree attainment and the health outcome. To avoid making this assumption, I estimated IPTW regression models as described above. The first point to make about the estimates of the effect of a college degree on depression and BMI is that neither estimate is statistically significant (Table 12). The effects are smaller, and have larger standard errors, but are directionally consistent. However, because the purpose of this study is to investigate genetic confounding rather than to estimate causal effects of a college degree on health, this is not a huge concern. We already know that for depression and BMI, including the propensity score as a covariate in the model captured any effect of genetic confounding. For self-rated health, the IPTW model can be more informative. In Model 1 in Table 12, graduating from college is significantly associated with a 0.159 ($0.160/\sigma_{y^*}$) standard deviation increase in self-rated health. After controlling for the three PGSs, this effect reduces to 0.149 ($0.153/\sigma_{y^*}$), which is a statistically insignificant decrease of 6%. Therefore, after estimating the effect of a college degree on self-rated health using an IPTW model, no evidence of genetic confounding remains. To summarize, while there was at least some evidence of genetic confounding in each education-health relationship under study, propensity score methods were able to remove the effect of genetic confounding in each case.

CONCLUSIONS AND DISCUSSION

Returning to the depiction of biological and mediated pleiotropy in Figure 1, the results here suggest that both processes are most likely behind the estimates of rG between education and health outcomes reported in other studies (Boardman et al. 2015; Bulik-Sullivan, Finucane, et al. 2015; Okbay et al. 2016; Wray et al. 2018). While this paper was more concerned with genetic confounding and therefore biological pleiotropy, there was evidence of mediated pleiotropy in each model. In fact, there was stronger support for mediated pleiotropy across the three health outcomes. College degree attainment always mediated the relationship between the education PGS and the health outcome, whether it was depression, BMI, or self-rated health.

The strongest evidence for genetic confounding emerged when examining the relationship between a college degree and BMI and between a college degree and self-rated health, and there was weaker evidence for genetic confounding of the college degree-depression relationship. Based on estimates of the proportion of the college degree effect that was confounded by common genetic factors, failing to adequately control for genetic endowments could result in slight overestimates of the effect of a college degree on depression, somewhat larger overestimates of the effect of a college degree on self-rated health, and substantial overestimates of the effect of a college degree on BMI.

PGSs have gained popularity recently in the social sciences partially because they are the first individual level measure of genome-wide genetic influence that can be easily incorporated into traditional social science analyses. In fact, one of their advertised benefits is that they can be used to adjust for previously unmeasured genetic confounding (Freese 2018). The PGSs clearly

were useful controls for genetic confounding in the initial models run without propensity scores. However, propensity score methods were able to completely block the effects of genetic confounding for each health outcome. It is important to note that the propensity scores employed were generated using earlier measures of all the health outcomes examined in this paper. If researchers do not have access to longitudinal data or a wide-raging set of controls as applied here, including PGSs in analyses of the education-health relationship will be effective. Even so, these results cast some doubt on the idea that PGSs have much more to offer as control variables over and above other covariates common to cohort studies in the social sciences. Unless they become much more predictive, they will certainly not serve as a magic bullet.

An important conclusion to draw from these findings is that the impact of genetic confounding could be much weaker than one might think by looking at estimates of rG alone. Past research has reported estimates of rG between educational attainment and self-rated health of 0.59 (Harris et al. 2017), educational attainment and BMI of -0.26 for BMI (Okbay et al. 2016), and college completion and depression of -0.17 (Wray et al. 2018). In this paper, the estimated proportion of the effect of a college degree confounded by genetic factors was 6% for self-rated health and 14% for BMI, with weak evidence of genetic confounding for depression. These results emphasize the inability of rG estimates alone to determine when we should be concerned about genetic confounding.

While these findings are important for future research on the relationship between education and health, future research should also consider the pathways through which this confounding effect of common genetic factors works for the relationship between college completion and both BMI and self-rated health. The PGSs used in this analysis are primarily predictive tools, and researchers are still trying to understand the causal pathways between the

genetic variants they represent and the phenotypes they predict. The common genetic variants that affect both education and BMI certainly do not do so directly. They likely operate through several proximate determinants.

Like all studies, this study has limitations. PGSs in general, including the ones in this analysis, suffer from measurement error, and do not represent the entire effect of the genome on any phenotype, much less all common genetic effects on two separate traits such as those studied here. A PGS for self-rated health was also not available. While PGSs will likely become more predictive in the future, they will likely not improve to the point of measuring all important genetic effects. Because the PGSs used in this analysis cannot capture all genetic effects on college degree attainment or the health outcomes, the results could understate the role of genetic confounding.

Furthermore, the utility of propensity scores to control for confounding is determined by the covariates used to generate the propensity scores. There are most likely other important unmeasured confounders that were not included in the propensity score generation. Many of the covariates used to generate the propensity score are affected by measurement error, and it is not clear how this might affect the accuracy of the college degree effect estimate. However, the goal of this paper was not to estimate a causal effect on health, but to test for genetic confounding.

FIGURES AND TABLES

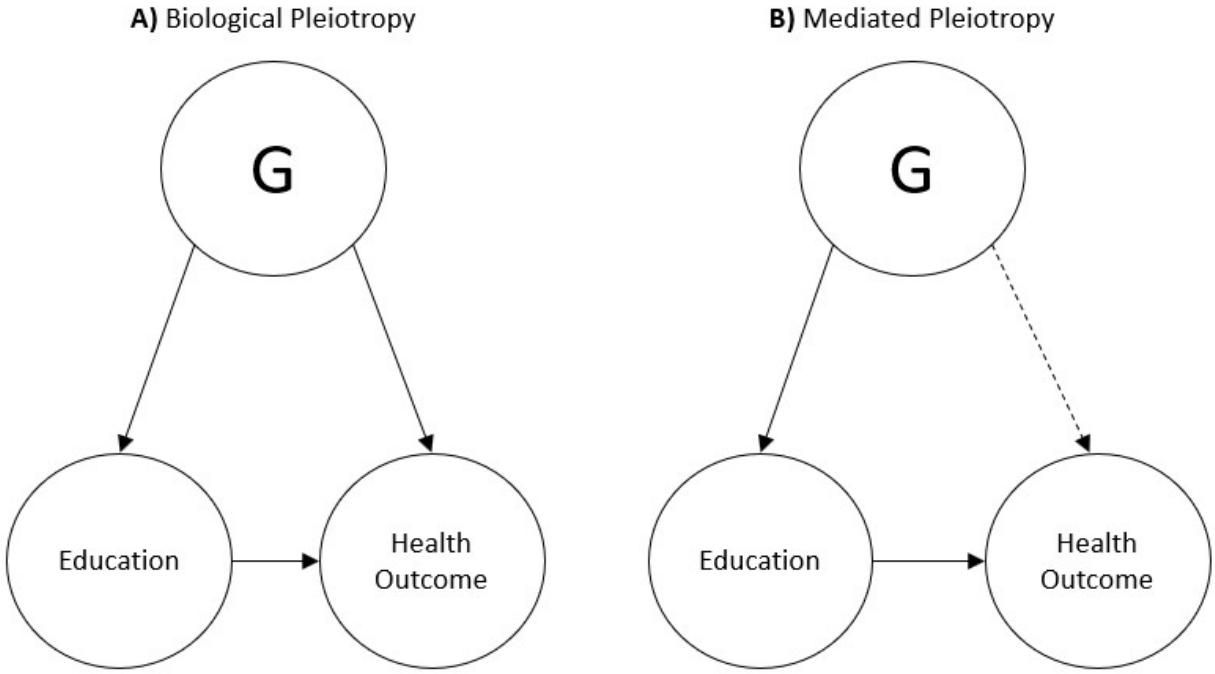

Figure 1. Conceptual diagrams representing two types of pleiotropy
Note: "G" represents additive genetic influence on both education and the health outcome of interest.

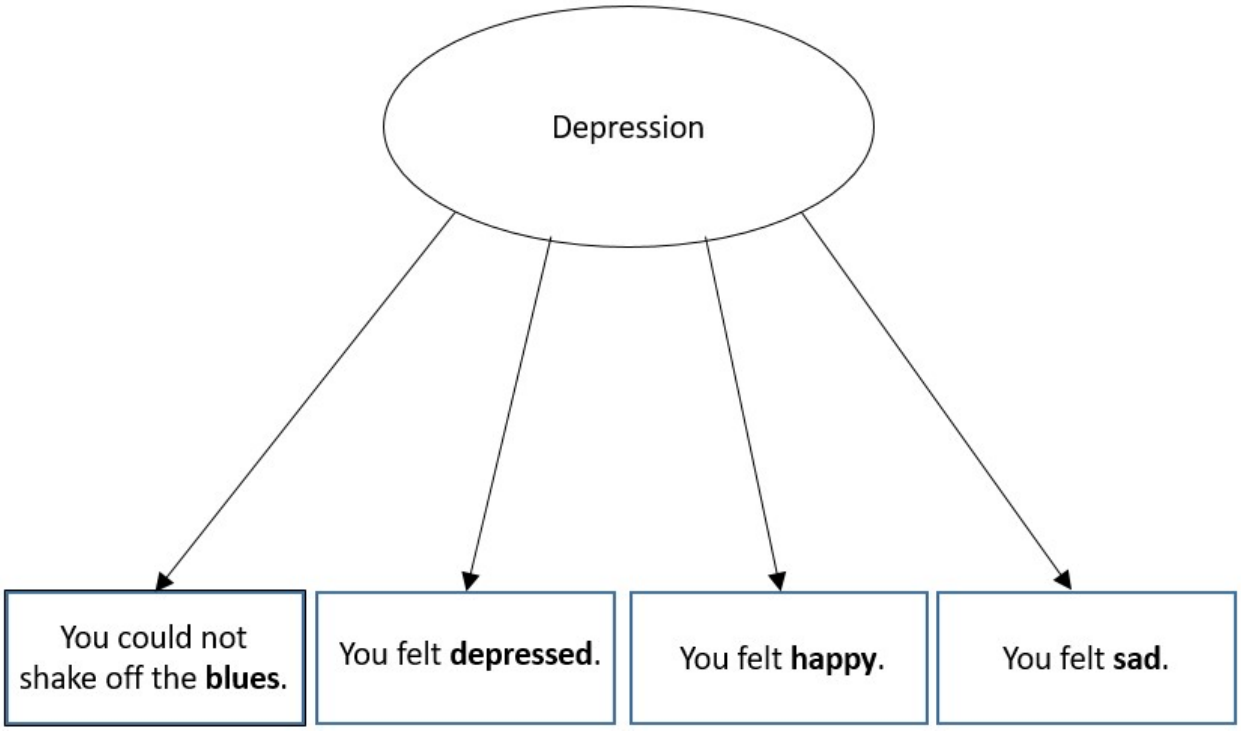

Figure 2. Measurement model of depression

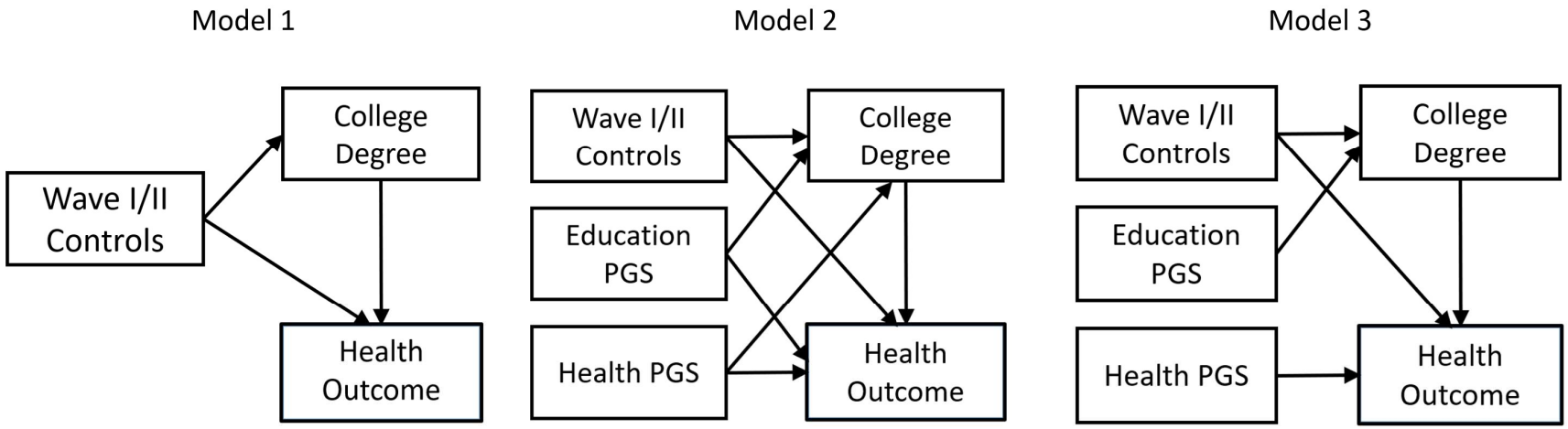

Figure 3. Structural equation models estimated for each health outcome.
Notes: Exogenous observed variables can correlate. Depression is modeled as a latent variable according to the measurement model in Figure 2.

Table 1. Summary Statistics

	Count of non-missing	Mean/%	S.D.	Min	Max
Focal Variables					
BMI	5369	28.64	7.36	14.40	98.00
Self-rated health	5629				
Poor		1.15%			
Fair		6.98%			
Good		32.26%			
Very good		40.93%			
Excellent		18.67%			
Depression indicators					
Depressed	5629				
Never or rarely		71.10%			
Sometimes		22.42%			
A lot of the time		4.26%			
Most of the time		2.22%			
Blues	5628				
Never or rarely		78.64%			
Sometimes		16.22%			
A lot of the time		3.62%			
Most of the time		1.51%			
Happy (reverse-coded)	5628				
Most of the time		41.17%			
A lot of the time		38.73%			
Sometimes		18.60%			
Never or rarely		1.49%			
Sad	5629				
Never or rarely		53.28%			
Sometimes		40.31%			
A lot of the time		4.81%			
Most of the time		1.60%			
College degree	5629	32.62%			
Education PGS	5629	0.00	1.00	-4.20	3.40
Depression PGS	5629	0.00	1.00	-3.88	3.63
BMI PGS	5629	0.00	1.00	-3.59	3.87
Variables for Propensity Scores					
Mom professional	5620	27.81%			
Dad professional	5611	78.27%			
No health insurance	5068	8.74%			
Income-to-needs ratio	5629				
Below 100%		8.56%			
100 – < 200%		17.25%			

200 – < 300%		19.45%			
300 – < 400%		15.24%			
400%+		21.32%			
Missing		18.17%			
Parent average years of edu.	5564	13.46	2.18	0.00	18.00
Female	5629	52.83%			
Age	5629	28.43	1.77	24.00	34.00
Vocabulary score	5390	104.98	11.79	18.00	138.00
Disabled	5629	3.41%			
Household smoker	5064	49.74%			
Freq. of parent heavy drinking	5057	1.26	0.82	1.00	6.00
Parent receives public assistance	4990	21.54%			
Parent smoker	5570	70.50%			
U.S. born	5629	98.99%			
Social control scale	5029	3.93	0.92	1.00	5.00
Parent-child closeness scale	5523	-0.04	0.97	-1.22	4.44
Parent disappointment if child does not graduate college	5102				
Very disappointed		35.91%			
Somewhat disappointed		45.16%			
Not disappointed		18.93%			
Household size	5623	4.35	1.33	1.00	15.00
Ever repeated grade	5626	18.86%			
Ever suspended	5627	21.98%			
Ever expelled	5623	2.69%			
Ever truant	5520	26.70%			
Standardized scale of grades	5505	-0.12	1.02	-1.69	3.82
School integration scale	5524	1.46	0.71	0.00	4.00
Getting along with teachers scale	5524	0.87	0.94	0.00	4.00
Problem with attention scale	5523	1.31	1.02	0.00	4.00
Problems with homework scale	5524	1.23	1.07	0.00	4.00
Getting along with students scale	5523	0.90	0.95	0.00	4.00
College expectations scale	5613	4.14	1.18	1.00	5.00
Desire to attend college scale	5614	4.39	1.08	1.00	5.00
Expectations to live to 35 scale	5608	4.49	0.78	1.00	5.00
Expectations killed by 21 scale	5618	-0.02	0.83	-0.83	4.14
Protective factors scale	5615	0.02	0.60	-2.79	1.13
Depressive symptoms scale	5627	-0.10	0.96	-1.42	5.60
Ever had sex	5596	35.67%			
Self-rated health	5627	3.88	0.89	1.00	5.00
How often missed school	5611	0.40	0.64	0.00	4.00
Smoking status	5581				

Daily smoker		13.33%			
Former smoker		3.92%			
Non-smoker		73.70%			
Infrequent smoker		9.05%			
Number of close friends that smoke	5612	0.16	0.96	-0.82	2.05
BMI	5506	22.31	4.38	12.01	54.23
Alcohol Consumption	5563				
Nondrinker		48.28%			
Usually has one drink		10.28%			
Usually has two drinks		7.68%			
Usually has 3+ drinks		33.76%			
Days in past year drunk	5590	0.72	1.26	0.00	6.00
Number of close friends that drink	5570	1.20	1.20	0.00	3.00
Physical activities in last week	5628	5.51	3.71	0.00	15.00
Visited dentist within last year	5619	72.98%			
Vegetable consumption	5626				
None		29.38%			
Once		39.35%			
Twice		31.27%			
Sweet snack consumption	5628				
None		45.24%			
Once		34.19%			
Twice		20.58%			
How often wears seatbelt	5628	3.13	1.15	0.00	4.00
Usually gets enough sleep	5626	73.76%			
Hours of screen time	5616	20.60	19.02	0.00	168.00
Delinquent behaviors scale	5613	-0.05	0.93	-0.73	8.89
Religious attendance scale	5541	1.63	1.21	0.00	3.00
Religious importance scale	5537	1.92	1.08	0.00	3.00
Neighborhood quality scale	5615	-0.15	0.96	-5.36	6.54

Table 2. Correlation matrix of primary variables of interest**.** N = 5,629

	Education PGS	BMI PGS	Depression PGS	College Degree	Depression	BMI	Self-rated health
Education PGS	1						
BMI PGS	-0.182***	1					
Depression PGS	-0.140***	-0.033	1				
College Degree	0.319***	-0.129***	-0.122***	1			
Depression	-0.072***	0.037	0.065**	-0.231***	1		
BMI	-0.063***	0.255***	0.018	-0.218***	0.000	1	
Self-rated health	0.117***	-0.129***	-0.058**	0.353***	-0.328***	-0.305***	1

Note: For all correlations involving BMI, N=5,425 after exclusion of individuals who were or might have been pregnant at Wave IV.

* p<0.05 ** p<0.01 *** p<0.001

Table 3. Fit statistics for depression models. N = 5,629

Fit statistics	**CFA**	**Model 1**	**Model 2**	**Model 3**	**Prop. Model 1**	**Prop. Model 2**	**Prop. Model 3**
Chi-squared	0.367	164.375	65.218	74.783	43.021	26.056	21.731
df	2	66	62	64	18	14	16
TLI	1	0.986	1	0.998	0.998	0.998	0.999
CFI	1	0.991	1	0.999	0.998	0.999	1
RMSEA	0	0.016	0.003	0.005	0.016	0.012	0.008
WRMR	0.087	1.061	0.540	0.611	1.021	0.598	0.621
Raftery's BIC	-16.904	-405.580	-470.1946	-477.901	-112.421	-94.844	-116.440

Note: All fit statistics are the means over the ten imputed datasets. Mplus does not provide a significance test for the chi-sqared statistic in this case. CFA - confirmatory factor analysis; Model 1 constrains the effect of the PGSs on both college degree attainment and depression to zero. Model 2 freely estimates the effect of the PGSs. Model 3 constraints the effect of the education PGS on depression and the effect of the depression PGS on college degree to zero. Prop. Models 1-3 follow the same pattern as Models 1-3, but they replace the observed covariates with a propensity score as discussed in the text.

Table 4. Structural Equation Models Predicting Depression.

	Model 1		Model 2		Model 3	
Dependent Variable	College Degree	Depression	College Degree	Depression	College Degree	Depression
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		-1.352*** (0.211)		-1.301*** (0.211)		-1.321*** (0.212)
Polygenic scores						
Education PGS			0.370*** (0.046)	-0.054 (0.089)	0.388*** (0.046)	
Depression PGS			-0.146*** (0.042)	0.197* (0.097)		0.203* (0.097)
Parent average years of edu.	0.370*** (0.029)	-0.076 (0.062)	0.350*** (0.029)	-0.073 (0.063)	0.349*** (0.029)	-0.076 (0.062)
Mom professional	0.236 (0.124)	-0.058 (0.233)	0.221 (0.124)	-0.052 (0.235)	0.226 (0.123)	-0.054 (0.235)
Dad professional	0.396** (0.125)	-0.566* (0.256)	0.387** (0.123)	-0.565* (0.256)	0.394*** (0.123)	-0.565* (0.257)
Income-to-needs ratio (400%+)						
Below 100%	-1.068*** (0.260)	-0.337 (0.453)	-1.018*** (0.265)	-0.346 (0.458)	-1.015*** (0.264)	-0.338 (0.456)
100 – < 200%	-0.974*** (0.162)	-0.213 (0.273)	-0.935*** (0.169)	-0.210 (0.277)	-0.931*** (0.165)	-0.207 (0.277)
200 – < 300%	-0.454*** (0.143)	-0.283 (0.286)	-0.459*** (0.143)	-0.264 (0.287)	-0.451** (0.144)	-0.265 (0.287)
300 – < 400%	-0.158 (0.142)	0.286 (0.291)	-0.223 (0.148)	0.310 (0.292)	-0.209 (0.151)	0.303 (0.290)
Missing	-0.217 (0.123)	-0.335 (0.309)	-0.229 (0.136)	-0.322 (0.308)	-0.216 (0.134)	-0.323 (0.308)
No health insurance	-0.586** (0.192)	0.218 (0.407)	-0.619** (0.196)	0.224 (0.408)	-0.615** (0.194)	0.220 (0.408)
Household smoker	-0.497*** (0.120)	0.231 (0.230)	-0.448*** (0.115)	0.219 (0.232)	-0.447*** (0.116)	0.225 (0.232)
Parent smoker	0.059 (0.102)	0.189 (0.218)	0.039 (0.102)	0.192 (0.218)	0.038 (0.102)	0.190 (0.219)
Freq. of parent heavy drinking	-0.021 (0.057)	0.070 (0.095)	-0.021 (0.058)	0.064 (0.095)	-0.026 (0.058)	0.063 (0.095)
Parent receives public assistance	-0.324* (0.156)	0.242 (0.271)	-0.300 (0.159)	0.234 (0.270)	-0.297 (0.157)	0.238 (0.272)
U.S. born	0.146 (0.430)	-1.176* (0.590)	0.209 (0.448)	-1.227* (0.584)	0.169 (0.443)	-1.228* (0.586)
Household size	0.035 (0.044)	0.051 (0.062)	0.025 (0.043)	0.053 (0.061)	0.025 (0.042)	0.052 (0.061)
Age	-0.002 (0.029)	0.032 (0.056)	-0.005 (0.030)	0.033 (0.056)	-0.005 (0.030)	0.033 (0.056)
Female	0.485*** (0.107)	0.593*** (0.185)	0.509*** (0.106)	0.579** (0.187)	0.498*** (0.104)	0.582** (0.186)
N	5629		5629		5629	

Note: Models are adjusted for the complex survey design of Add Health. Equations predicting college degree use a logit link function.

*** p<0.001 ** p<0.01 * p<0.05

Table 5. Fit statistics for BMI models. N = 5,425

Fit statistics	**Model 1**	**Model 2**	**Model 3**	**Prop. Model 1**	**Prop. Model 2**	**Prop. Model 3**
Chi-squared	269.619	-	9.068	199.090	-	4.531
df	4	-	2	4	-	2
TLI	-5.378	-	0.660	0.706	-	0.992
CFI	0.346	-	0.983	0.832	-	0.998
RMSEA	0.111	-	0.026	0.095	-	0.015
WRMR	2.578	-	0.459	4.522	-	0.672
Raftery's BIC	235.224	-	-8.130	164.695	-	-12.667

Note: All fit statistics are the means over the ten imputed datasets. Mplus does not provide a significance test for the chi-squared statistic in this case. Model 1 constrains the effect of the PGSs on both college degree attainment and BMI to zero. Model 2 freely estimates the effect of the PGSs, and is exactly identified, so no fit statistics are available. Model 3 constraints the effect of the education PGS on BMI and the effect of the BMI PGS on college degree to zero. Prop. Models 1-3 follow the same pattern as Models 1-3, but they replace the observed covariates with a propensity score as discussed in the text.

Table 6. Structural Equation Models Predicting BMI.

	Model 1		Model 2		Model 3	
Dependent Variable	College Degree	BMI	College Degree	BMI	College Degree	BMI
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		-1.652*** (0.276)		-1.425*** (0.297)		-1.366*** (0.295)
Polygenic scores						
Education PGS			0.357*** (0.046)	0.179 (0.143)	0.372*** (0.045)	
BMI PGS			-0.099* (0.042)	1.848*** (0.126)		1.821*** (0.123)
Parent average years of edu.	0.367*** (0.030)	-0.342*** (0.095)	0.347*** (0.029)	-0.324*** (0.095)	0.348*** (0.030)	-0.316*** (0.093)
Mom professional	0.244 (0.127)	0.673 (0.345)	0.230 (0.125)	0.655 (0.356)	0.231 (0.125)	0.661 (0.357)
Dad professional	0.408** (0.130)	0.207 (0.359)	0.412*** (0.128)	0.117 (0.340)	0.405** (0.129)	0.118 (0.339)
Income-to-needs ratio (400%+)						
Below 100%	-1.001*** (0.263)	1.505* (0.634)	-0.952*** (0.268)	1.495* (0.609)	-0.950*** (0.268)	1.467* (0.611)
100 – < 200%	-0.981*** (0.165)	0.848 (0.454)	-0.934*** (0.169)	0.747 (0.440)	-0.939*** (0.168)	0.737 (0.442)
200 – < 300%	-0.482*** (0.148)	1.215** (0.422)	-0.473** (0.150)	1.145** (0.419)	-0.477*** (0.149)	1.149** (0.418)
300 – < 400%	-0.176 (0.143)	0.456 (0.518)	-0.209 (0.150)	0.291 (0.501)	-0.220 (0.150)	0.315 (0.496)
Missing	-0.232 (0.128)	0.756 (0.438)	-0.227 (0.140)	0.700 (0.425)	-0.231 (0.139)	0.700 (0.425)
No health insurance	-0.578** (0.195)	-0.817 (0.474)	-0.607** (0.198)	-0.695 (0.483)	-0.609** (0.197)	-0.687 (0.482)
Household smoker	-0.512*** (0.125)	0.345 (0.391)	-0.461*** (0.120)	0.300 (0.371)	-0.462*** (0.120)	0.279 (0.376)
Parent smoker	0.061 (0.105)	0.228 (0.319)	0.046 (0.105)	0.135 (0.314)	0.041 (0.104)	0.141 (0.314)
Freq. of parent heavy drinking	-0.016 (0.060)	-0.329* (0.167)	-0.022 (0.061)	-0.308 (0.158)	-0.021 (0.060)	-0.306 (0.158)
Parent receives public assistance	-0.311 (0.162)	-0.056 (0.393)	-0.279 (0.164)	-0.133 (0.397)	-0.285 (0.163)	-0.143 (0.398)
U.S. born	0.108 (0.429)	1.523 (1.087)	0.143 (0.440)	1.410 (1.152)	0.132 (0.441)	1.402 (1.164)
Household size	0.032 (0.044)	-0.232* (0.105)	0.021 (0.043)	-0.223* (0.102)	0.022 (0.043)	-0.220* (0.102)
Age	-0.001 (0.030)	0.173* (0.070)	-0.004 (0.031)	0.162* (0.069)	-0.004 (0.031)	0.164* (0.069)
Female	0.459*** (0.113)	-0.273 (0.262)	0.479*** (0.110)	-0.391 (0.267)	0.473*** (0.110)	-0.399 (0.266)
N	5425		5425		5425	

Note: Models are adjusted for the complex survey design of Add Health. Equations predicting college degree use a logit link function.
*** p<0.001 ** p<0.01 * p<0.05

Table 7. Fit statistics for self-rated health models. N = 5,629

Fit statistics	**Model 1**	**Model 2**	**Model 3**	**Prop. Model 1**	**Prop. Model 2**	**Prop. Model 3**
Chi-squared	121.219	-	18.149	50.494	-	3.449
df	6	-	3	6	-	3
TLI	-0.908	-	0.498	0.946	-	0.999
CFI	0.721	-	0.963	0.964	-	1
RMSEA	0.058	-	0.030	0.036	-	0.005
WRMR	1.652	-	0.636	1.943	-	0.499
Raftery's BIC	69.405	-	-7.758	-1.320	-	-22.458

Note: All fit statistics are the means over the ten imputed datasets. Mplus does not provide a significance test for the chi-squared statistic in this case. Model 1 constrains the effect of the PGSs on both college degree attainment and self-rated health to zero. Model 2 freely estimates the effect of the PGSs, and is exactly identified, so no fit statistics are available. Model 3 constraints the effect of the education PGS on self-rated health and the effect of the BMI and depression PGSs on college degree to zero. Prop. Models 1-3 follow the same pattern as Models 1-3, but they replace the observed covariates with a propensity score as discussed in the text.

Table 8. Structural Equation Models Predicting Self-rated Health.

	Model 1		Model 2		Model 3	
Dependent Variable	College Degree	SRH	College Degree	SRH	College Degree	SRH
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		0.472*** (0.041)		0.445*** (0.042)		0.451*** (0.041)
Polygenic scores						
Education PGS			0.350*** (0.046)	0.017 (0.020)	0.388*** (0.046)	
BMI PGS			-0.124** (0.043)	-0.101*** (0.017)		-0.104*** (0.017)
Depression PGS			-0.154*** (0.043)	-0.033 (0.021)		-0.035 (0.021)
Parent average years of edu.	0.370*** (0.029)	0.025* (0.011)	0.349*** (0.029)	0.023* (0.011)	0.349*** (0.029)	0.024* (0.011)
Mom professional	0.236 (0.124)	-0.026 (0.045)	0.219 (0.124)	-0.028 (0.045)	0.226 (0.123)	-0.027 (0.045)
Dad professional	0.396** (0.125)	0.069 (0.051)	0.396*** (0.123)	0.073 (0.050)	0.394*** (0.123)	0.073 (0.050)
Income-to-needs ratio (400%+)						
Below 100%	-1.068*** (0.260)	-0.051 (0.085)	-1.015*** (0.266)	-0.046 (0.084)	-1.015*** (0.264)	-0.049 (0.083)
100 – < 200%	-0.974*** (0.162)	-0.130 (0.066)	-0.928*** (0.169)	-0.125 (0.065)	-0.931*** (0.165)	-0.125 (0.065)
200 – < 300%	-0.454*** (0.143)	-0.061 (0.057)	-0.453** (0.144)	-0.060 (0.056)	-0.451** (0.144)	-0.060 (0.056)
300 – < 400%	-0.158 (0.142)	-0.063 (0.066)	-0.209 (0.147)	-0.060 (0.067)	-0.209 (0.151)	-0.057 (0.067)
Missing	-0.217 (0.123)	-0.032 (0.062)	-0.225 (0.137)	-0.032 (0.062)	-0.216 (0.134)	-0.032 (0.062)
No health insurance	-0.586** (0.192)	0.069 (0.070)	-0.618** (0.196)	0.059 (0.069)	-0.615** (0.194)	0.060 (0.069)
Household smoker	-0.497*** (0.120)	-0.181 (0.052)	-0.447*** (0.114)	-0.176*** (0.051)	-0.447*** (0.116)	-0.177*** (0.051)
Parent smoker	0.059 (0.102)	-0.029 (0.051)	0.046 (0.102)	-0.025 (0.051)	0.038 (0.102)	-0.025 (0.051)
Freq. of parent heavy drinking	-0.021 (0.057)	0.000 (0.022)	-0.023 (0.059)	-0.001 (0.022)	-0.026 (0.058)	-0.001 (0.022)
Parent receives public assistance	-0.324* (0.156)	-0.132 (0.052)	-0.292 (0.159)	-0.126* (0.053)	-0.297 (0.157)	-0.127* (0.052)
U.S. born	0.146 (0.430)	-0.034 (0.179)	0.227 (0.449)	-0.018 (0.185)	0.169 (0.443)	-0.019 (0.185)
Household size	0.035 (0.044)	0.007 (0.015)	0.023 (0.043)	0.006 (0.015)	0.025 (0.042)	0.007 (0.015)
Age	-0.002 (0.029)	-0.004 (0.012)	-0.006 (0.030)	-0.004 (0.012)	-0.005 (0.030)	-0.004 (0.012)
Female	0.485*** (0.107)	-0.028 (0.036)	0.515*** (0.106)	-0.019 (0.037)	0.498*** (0.104)	-0.019 (0.037)
N	5629		5629		5629	

Note: Models are adjusted for the complex survey design of Add Health. The equations predicting college degree use a logit link function and the equations predicting self-rated health use a probit link function.
*** p<0.001 ** p<0.01 * p<0.05

Table 9. Structural equation models with propensity scores predicting depression.

	Model 1		Model 2		Model 3	
Dependent Variable	College Degree	Depression	College Degree	Depression	College Degree	Depression
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		-0.632* (0.253)		-0.621* (0.254)		-0.621* (0.253)
Polygenic scores						
Education PGS			0.195*** (0.052)	0.001 (0.086)	0.201*** (0.050)	
Depression PGS			-0.052 (0.048)	0.189 (0.097)		0.189 (0.097)
Propensity score	6.277*** (0.197)	-2.293*** (0.452)	6.111 0.193	-2.243*** (0.445)	6.123*** (0.193)	-2.242*** (0.451)
N	5629		5629		5629	

Note: Models are adjusted for the complex survey design of Add Health. Equations predicting college degree use a logit link function.

*** p<0.001 ** p<0.01 * p<0.05

Table 10. Structural equation models with propensity scores predicting BMI.

	Model 1		Model 2		Model 3	
Dependent Variable	College Degree	BMI	College Degree	BMI	College Degree	BMI
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		-1.071** (0.358)		-0.964** (0.371)		-0.934* (0.369)
Polygenic scores						
Education PGS			0.173*** (0.051)	0.209 (0.134)	-2.948*** (0.594)	
BMI PGS			-0.064 (0.050)	1.845*** (0.128)		1.814*** (0.126)
Propensity score	6.269*** (0.204)	-3.294*** (0.609)	6.119*** (0.200)	-2.948*** (0.594)	6.126*** (0.200)	-2.779*** (0.586)
N	5425		5425		5425	

Note: Models are adjusted for the complex survey design of Add Health. Equations predicting college degree use a logit link function.
*** p<0.001 ** p<0.01 * p<0.05

Table 11. Structural equation models with propensity scores predicting self-rated health.

Dependent Variable	Model 1		Model 2		Model 3	
	College Degree	SRH	College Degree	SRH	College Degree	SRH
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
College degree		0.300*** (0.048)		0.290*** (0.049)		0.291*** (0.049)
Polygenic scores						
Education PGS			0.181*** (0.052)	0.007 (0.020)	0.201*** (0.050)	
BMI PGS			-0.079 (0.049)	-0.100*** (0.017)		-0.101*** (0.017)
Depression PGS			-0.058 (0.049)	-0.028 (0.023)		-0.029 0.022
Propensity score	6.277*** (0.197)	0.729*** (0.080)	6.101*** (0.194)	0.691*** (0.081)	6.123*** (0.193)	0.696*** (0.082)
N	5629		5629		5629	

Note: Models are adjusted for the complex survey design of Add Health. The equations predicting college degree use a logit link function and the equations predicting self-rated health use a probit link function.
*** p<0.001 ** p<0.01 * p<0.05

Table 12. Inverse probability of treatment weighted models

Dependent Variable	Depression - Coef. (SE)	BMI - Coef. (SE)	Self-rated health - Coef. (SE)	
			Model 1	Model 2
College degree	-0.239 (0.394)	-0.623 (0.451)	0.160* (0.069)	0.153* (0.069)
Polygenic scores				
Education PGS				0.055 (0.033)
BMI PGS				-0.203*** (0.043)
Depression PGS				-0.001 (0.037)
N	5628	5424	5628	5628

Note: Models are adjusted for the complex survey design of Add Health. The equation predicting self-rated health uses a probit link function.
*** p<0.001 ** p<0.01 * p<0.05

REFERENCES

Amin, Vikesh, Jere R. Behrman, and Hans-Peter Kohler. 2015. "Schooling Has Smaller or Insignificant Effects on Adult Health in the US Than Suggested by Cross-sectional Associations: New Estimates Using Relatively Large Samples of Identical Twins." *Social Science & Medicine* 127:181–189.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate behavioral research* 46(3):399–424.

Bailis, Daniel S., Alexander Segall, and Judith G. Chipperfield. 2003. "Two Views of Self-rated General Health Status." *Social Science & Medicine* 56(2):203–217.

Behrman, Jere R. et al. 2011. "Does More Schooling Reduce Hospitalization and Delay Mortality? New Evidence Based on Danish Twins." *Demography* 48(4):1347–1375.

Behrman, Jere R., Yanyan Xiong, and Junsen Zhang. 2015. "Cross-sectional Schooling-health Associations Misrepresented Causal Schooling Effects on Adult Health and Health-related Behaviors: Evidence from the Chinese Adults Twins Survey." *Social Science & Medicine* 127:190–197.

Belsky, Daniel W. et al. 2018. "Genetic Analysis of Social-class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences of the United States of America* 115(31):E7275–E7284.

Belsky, Daniel W., and Salomon Israel. 2014. "Integrating Genetics and Social Science: Genetic Risk Scores." *Biodemography and social biology* 60(2):137–155.

Boardman, Jason D., Benjamin W. Domingue, and Jonathan Daw. 2015. "What Can Genes Tell Us About the Relationship Between Education and Health?" *Social Science & Medicine* 127:171–180.

Boardman, Jason D., and Jason M. Fletcher. 2015. "To Cause or Not to Cause? That Is the Question, but Identical Twins Might Not Have All of the Answers." *Social Science & Medicine* 127:198–200.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Bollen, Kenneth A., Jeffrey J. Harden, Surajit Ray, and Jane Zavisca. 2014. "BIC and Alternative Bayesian Information Criteria in the Selection of Structural Equation Models." *Structural Equation Modeling: A Multidisciplinary Journal* 21(1):1–19.

Bollen, Kenneth A., and Robert Stine. 1990. "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability." *Sociological methodology* 20:115.

Braudt, David B., and Kathleen M. Harris. 2018. *Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 1*. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Retrieved August 14, 2018 (http://www.cpc.unc.edu/projects/addhealth/documentation/restricteduse/datasets/WaveIVPolygenicScoresPGS.zip).

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models." *Annual review of sociology* 44(1):39–54.

Bulik-Sullivan, Brendan, Hilary K. Finucane, et al. 2015. "An Atlas of Genetic Correlations Across Human Diseases and Traits." *Nature Genetics* 47(11):1236–1241.

Bulik-Sullivan, Brendan, Po-Ru Loh, et al. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-wide Association Studies." *Nature Genetics* 47(3):291–295.

Davies, Neil M., Matt Dickson, George Davey Smith, Gerard J. van den Berg, and Frank Windmeijer. 2018. "The Causal Effects of Education on Health Outcomes in the UK Biobank." *Nature human behaviour*.

Deaton, Angus. 2009. *Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development*. Cambridge, MA: National Bureau of Economic Research.

Dudbridge, Frank. 2013. "Power and Predictive Accuracy of Polygenic Risk Scores." *PLoS Genetics* 9(3):e1003348.

Fletcher, Jason M. 2015. "New Evidence of the Effects of Education on Health in the US: Compulsory Schooling Laws Revisited." *Social Science & Medicine* 127:101–107.

Freese, Jeremy. 2018. "The Arrival of Social Science Genomics." *Contemporary sociology* 47(5):524–536.

Gaydosh, Lauren, Daniel W. Belsky, Benjamin W. Domingue, Jason D. Boardman, and Kathleen Mullan Harris. 2018. “Father Absence and Accelerated Reproductive Development in Non-Hispanic White Women in the United States.” *Demography* 55(4):1245–1267.

Haberstick, Brett C. et al. 2010. “Stable Genes and Changing Environments: Body Mass Index Across Adolescence and Young Adulthood.” *Behavior Genetics* 40(4):495–504.

Harris, Sarah E. et al. 2017. “Molecular Genetic Contributions to Self-rated Health.” *International Journal of Epidemiology* 46(3):994–1009.

Hermann, Silke et al. 2011. “The Association of Education with Body Mass Index and Waist Circumference in the EPIC-PANACEA Study.” *BMC Public Health* 11:169.

Highland, Heather M., Christy L. Avery, Qing Duan, Yun Li, and Kathleen Mullan Harris. 2018. *Quality Control Analysis of' ' Add Health GWAS Data*. Chapel Hill, NC: Carolina Population Center. Retrieved August 15, 2018 (http://www.cpc.unc.edu/projects/addhealth/documentation/guides/copy_of_AH_GWAS_QC.pdf).

Hussey, Jon M. et al. 2015. “The Reliability of In-home Measures of Height and Weight in Large Cohort Studies: Evidence from Add Health.” *Demographic Research* 32:1081–1098.

Johnson, W., M. McGue, D. Gaist, J. W. Vaupel, and K. Christensen. 2002. “Frequency and Heritability of Depression Symptomatology in the Second Half of Life: Evidence from Danish Twins over 45.” *Psychological Medicine* 32(7):1175–1185.

Kane, Jennifer B., Kathleen Mullan Harris, S. Philip Morgan, and David K. Guilkey. 2018. “Pathways of Health and Human Capital from Adolescence into Young Adulthood.” *Social forces; a scientific medium of social study and interpretation* 96(3):949–976.

Lawrence, Elizabeth M. 2017. “Why Do College Graduates Behave More Healthfully Than Those Who Are Less Educated?” *Journal of Health and Social Behavior* 58(3):291–306.

Lee, James J. et al. 2018. “Gene Discovery and Polygenic Prediction from a Genome-wide Association Study of Educational Attainment in 1.1 Million Individuals.” *Nature Genetics* 50(8):1112–1121.

Leinonen, Raija et al. 2005. “Genetic Influences Underlying Self-rated Health in Older Female Twins.” *Journal of the American Geriatrics Society* 53(6):1002–1007.

Link, B. G., and J. Phelan. 1995. “Social Conditions as Fundamental Causes of Disease.” *Journal of Health and Social Behavior* Spec No:80–94.

Liu, Hexuan. 2019. “Genetic Architecture of Socioeconomic Outcomes: Educational Attainment, Occupational Status, and Wealth.” *Social science research* 82:137–147.

Lleras-Muney, A. 2005. “The Relationship Between Education and Adult Mortality in the United States.” *The Review of economic studies* 72:189–221.

Locke, Adam E. et al. 2015. “Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology.” *Nature* 518(7538):197–206.

MacKinnon, D. P., J. L. Krull, and C. M. Lockwood. 2000. “Equivalence of the Mediation, Confounding and Suppression Effect.” *Prevention Science* 1(4):173–181.

Mazumder, Bhashkar. 2010. “Erratum: Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws.” *SSRN Electronic Journal*.

Miech, Richard Allen, and Michael J. Shanahan. 2000. “Socioeconomic Status and Depression over the Life Course.” *Journal of Health and Social Behavior* 41(2):162.

Mirowsky, John, and Catherine E. Ross. 2008. “Education and Self-Rated Health.” R*esearch on aging* 30(1):93–122.

Montez, Jennifer Karas, Robert A. Hummer, and Mark D. Hayward. 2012. “Educational Attainment and Adult Mortality in the United States: a Systematic Analysis of Functional Form.” *Demography* 49(1):315–336.

Muthen, Linda, and Bengt Muthen. n.d. “Mplus Discussion - Multiple Imputation.” *Mplus*. Retrieved February 26, 2020 (http://www.statmodel.com/discussion/messages/22/381.html).

Okbay, Aysu et al. 2016. “Genome-wide Association Study Identifies 74 Loci Associated with Educational Attainment.” *Nature* 533(7604):539–542.

Perreira, Krista M., Natalia Deeb-Sossa, Kathleen Mullan Harris, and Kenneth Bollen. 2005. “What Are We Measuring? An Evaluation of the CES-D Across Race/Ethnicity and Immigrant Generation.” *Social Forces* 83(4):1567–1601.

Phelan, Jo C., Bruce G. Link, and Parisa Tehranifar. 2010. “Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications.” *Journal of Health and Social Behavior* 51 Suppl:S28–40.

Price, Alkes L. et al. 2006. “Principal Components Analysis Corrects for Stratification in Genome-wide Association Studies.” *Nature Genetics* 38(8):904–909.

Radloff, L. S. 1977. “The CES-D Scale: A Self-report Depression Scale for Research in the General Population.” *Applied psychological measurement* 1(3):385–401.

Raftery, Adrian E. 1995. “Bayesian Model Selection in Social Research.” *Sociological methodology* 25:111.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70(1):41.

Ross, Catherine E., and Chia-ling Wu. 1995. “The Links Between Education and Health.” *American Sociological Review* 60(5):719.

Selzam, Saskia et al. 2019. “Comparing Within- and Between-Family Polygenic Score Prediction.” *American Journal of Human Genetics* 105(2):351–363.

Sobel, Michael E. 1982. “Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models.” *Sociological methodology* 13:290.

Solovieff, Nadia, Chris Cotsapas, Phil H. Lee, Shaun M. Purcell, and Jordan W. Smoller. 2013. “Pleiotropy in Complex Traits: Challenges and Strategies.” *Nature Reviews. Genetics* 14(7):483–495.

Tropf, Felix C. et al. 2017. “Hidden Heritability Due to Heterogeneity Across Seven Populations.” *Nature human behaviour* 1(10):757–765.

Wedow, Robbee et al. 2018. “Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects.” *American Sociological Review* 83(4):802–832.

Wertz, Jasmin et al. 2019. “Using DNA from Mothers and Children to Study Parental Investment in Children’s Educational Attainment.” *Child Development*.

Wray, Naomi R. et al. 2018. “Genome-wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression.” *Nature Genetics* 50(5):668–681.

Young, Alexander I., Stefania Benonisdottir, Molly Przeworski, and Augustine Kong. 2019. “Deconstructing the Sources of Genotype-phenotype Associations in Humans.” *Science* 365(6460):1396–1400.

Zheng, Hui. 2017. “Why Does College Education Matter? Unveiling the Contributions of Selection Factors.” *Social science research* 68:59–73.

Zheng, Jie et al. 2017. “LD Hub: a Centralized Database and Web Interface to Perform LD Score Regression That Maximizes the Potential of Summary Level GWAS Data for SNP Heritability and Genetic Correlation Analysis.” *Bioinformatics* 33(2):272–279.